SCALING UP STRATEGIES FOR CHANGE

Truus Dekker Els Feijs

University of Utrecht

This research report builds on the results of the CATCH project[1](Webb and Dekker, 2002). The CATCH project is an effort to apply and scale up previous results and design a professional development program to bring about fundamental changes in teachers' instruction to support change in teachers' formative assessment practices. In this report we focus on two of the research questions: "What professional development materials support improved formative assessment across a wide range of schools?" and "How do teachers' conceptions and assessment practices change as a result of their participation in this development program?" Initial results show that a change in teachers' attitude towards formative assessment occurred. Moreover, changes in instructional assessment were easier to implement for teachers than ideas related to the pyramid model.

INTRODUCTION

When in The Netherlands a new mathematics exam program[2] was introduced in 1992, the possibility to change the mandatory central examinations at the same time proved to be of great importance. For teachers as well as curriculum designers the format of the questions (no more multiple choice questions), posed within a context, as well as the scoring guide that showed different strategies and sometimes even different possible (correct) answers, set an example for the daily classroom practice. As far as we know, no formal research has been conducted in The Netherlands to the extent of influence of central examinations on the teaching and learning process preceding it. Results from the Research in Assessment Practices[3] (RAP) project, preceding the CATCH project, showed that many teachers have limited understanding of formative assessment practices and, thus, provide students with incomplete information about their progress (Romberg, 1999). When teachers learn to utilize formative assessment practices in their classrooms as a consequence of appropriate professional development, there are positive effects on student learning and achievement (Black &William, 1998). When teachers retained conventional assessment techniques, especially focused on assessing basic skills, they also paid little attention to different strategies used by their students to solve a problem, to classroom discussions and, in general, to "teaching and learning with understanding", even when they used reform curricula.

It is on the basis of these past experiences that we assumed that changing the assessment practices of teachers already using reform curricula could play an important part to

---

[1] CATCH Classroom Assessment as a basis for Teacher Change. Funded by the Office of Educational Research and Improvement (OERI PR/award R305A60007) 2000-2003

[2] for students of vbo (vocational) and mavo (theoretical level, leading to vocational education at middle level), central examinations are held at the age of 16.

[3] RAP was one of the projects of the National Center for Improving Student Learning and Achievement in Mathematics and Science (NCISLA), funded by OERI, 305A60007-98,1996-2000

enhance “teaching and assessing for understanding”. By posing questions and using test items that evoke mathematical reasoning and generalizing and by asking students to choose their own mathematical tools to solve a problem, it becomes clear what students are able to do instead of which facts, standard algorithms and definitions they know. If the process is as important as the product, if strategies used by students are important, it becomes important for the teacher to listen carefully to what students say in class and assess student work more closely. Teachers use the information gathered this way to guide instruction. Textbooks do not always provide good problems. So the next step for teachers to be taken is to adapt questions posed in the curriculum or to enhance their own assessment problems. Then the design of balanced assessments is being discussed, using problems at different competency levels. Assessing becomes a continuous process, an integrated part of the teaching and learning process instead of something from outside, interrupting this process but nevertheless inevitable. We feel that this results in a more student centered instructional environment, more likely to improve student achievement (Bransford, Brown & Cocking, 1999).

PROFESSIONAL DEVELOPMENT: MATERIALS AND STRATEGY

The ideas on which the CATCH project, meant for mathematics teachers at middle grades, was based and which served as a starting point for the design of professional development materials find their origin in the work of the Freudenthal Institute. The ideas are based on studies on assessment by researchers at the Freudenthal Institute, (de Lange 1987, Van den Heuvel P, 2000). They place assessment as an essential part of the teaching learning process and emphasize assessing for understanding. An essential part of the principles underpinning the assessment principles is the distinction of mathematical competencies at three levels (De Lange 1996). These ideas were further expanded into the so-called pyramid model (Boertien en Verhage 1993,Verhage en de Lange 1995, de Lange 1999) when designing a National Option for the TIMSS[4] study (Kuiper, 2000).

These ideas, incorporated in the Framework for Classroom Assessment in Mathematics (de Lange 1999), that forms the theoretical framework for the CATCH project, are aligned by those of the OECD[5] Programme for International Student Assessment (PISA), (OECD, 1999).

In the aforementioned model three levels of mathematical competency are discerned:

Level 1: Reproduction, procedures, concepts and definitions;

Level 2: Connections and integration for problem solving;

Level 3: Mathematization, mathematical thinking and reasoning, generalization and insight.

A course for teachers as an elaboration of the framework was developed: *Great Assessment Problems (GAP)* (Dekker & Querelle, 2002). An assessment tool, *AssessMath!* (Cappo, De Lange & Romberg, 1999), developed earlier was also used.

---

[4] Third International Mathematics and Science Study

[5] Organisation for Economic Co-operation and Development

Teachers taking part in the project were continuously asked to give feedback on the materials being developed.

The strategy used in the project is based on the "train the trainer" model. The initial seminar, a three-day conference in Utrecht for about 20 teachers as well as administrators from two different school districts, was meant to be used by the CATCH lead teachers as a model for their own CATCH summer institutes. The Dutch CATCH team consequently developed a model seminar, consisting of different elements taken from these summer institutes. We want to take teachers seriously and value their experience, thus neither work "top down" nor "bottom up" but both ways at the same time. During the summer institutes CATCH lead teachers in the two districts used examples from their own students and from their own classrooms.

We expected the teachers taking part in the project to go through the following steps in their professional development:

1. During the initial CATCH seminar, lead teachers and administrators critique and develop greater understanding of existing assessment instruments using resources such as *Great Assessment Problems* and *AssessMath!* together with colleagues from their own district and/or school;
2. Teachers select and adapt assessment instruments for their own use with students and report about results during monthly and ad hoc meetings with colleagues;
3. In considering and using these instruments, teachers examine the role and function of assessment instruments versus the desired learning outcomes and the potential for positive feedback;
4. Scoring and grading of student work is used to provide insight in student (mis)conceptions that guide instruction;
5. Teachers implement design principles for classroom assessment and learn how a series of items can be constructed to design balanced assessments that reflect a hypothetical assessment trajectory. This provides students with the opportunity to demonstrate the full range of mathematical competencies including making mathematical arguments, non-routine problem solving, developing their own models and inventing new strategies;
6. Teachers explore assessment opportunities embedded within instructional contexts, learn how to balance the use of formal and instructional assessment, and examine the relationship between classroom assessment and student achievement on external assessments;
7. Teachers inform their colleagues during successive summer institutes, thus helping ideas and outcomes of the CATCH project to "travel" to new classrooms, schools and districts.

While passing through this trajectory we expected teachers to (a) recognize problems at different competency levels, (b) use or even design more higher level problems in their own assessments, (c) understand and use instructionally embedded assessment and (d) use a more varied set of assessment instruments in general.

It has been amazing, even to the members of the research team of the project, both in The Netherlands and the US, how fast ideas of the CATCH project "traveled". The two year project has not yet come to an end but in district A, a small urban/suburban district, serving over 3,000 students predominantly European American (85%) with

approximately 30% free or reduced-cost meals, members of the CATCH team have gained visibility and assumed greater responsibility in school and district leadership in enacting changes in use of curricula and fostering greater consistency in classroom assessment practices at each grade level. CATCH lead teachers in this district noted that the summer institute was a decisive turning point in re-directing colleagues to enjoin in collaborative decision making toward modifying instructional resources and improving classroom assessment instruments during monthly grade-level meetings.

In district B, a large urban district in Eastern United States, where middle grade teachers work with predominantly African American and Hispanic students, with more than 75% receiving free or reduced-cost meals, "traveling" has taken place to an even greater extent. The district is now making plans to expand the CATCH program to K-12 teachers of mathematics as well as science. Lead teachers in this district felt the pyramid model could also be adapted for other grades and other subjects.

ANALYSIS OF QUALITATIVE DATA

Qualitative data[6] for this project from 14 lead teachers [7] was gathered through an initial survey, three rounds of classroom observations and teacher interviews, and through collecting assessment portfolios. Observation protocols were adapted from instruments used by Horizon Research[8]. An exit survey for participating lead teachers will follow by the end of the project. Analysis of the interviews is being conducted, using the Multiple Episode Protocol Analysis (MEPA)[9] program. Interjudge agreement is secured by having analyzed and categorized the data independently by at least two members of the CATCH research team. Since this is only a relatively small study, the results of the second round of interviews were used to assess and adapt the codes used. Observations and assessment portfolios will be used to analyze individual teacher development and to validate the research implications derived from the interviews. The research questions, posed in the project were:

1. What professional development materials will be required to disseminate principles for improving formative assessment across a wide range of schools?
2. What support do school personnel and teachers in various school contexts, who are adapting these principles to local conditions, need to ensure that changes in formative assessment are sustained?
3. How do teachers make decisions about what assessment instruments to use, when to use them, and which reasons motivate their choices?
4. How do teachers' assessment practices change as a result of their participation in this professional development program?

---

[6] quantitative data, based on standardized test scores of students taught by CATCH teachers, are also being collected and will be analyzed later.

[7] 6 teachers from district A, 8 from district B.

[8] see http://www.horizon-research.com/LSC/manual/0203/existing.php#6

[9] MEPA was designed by Gijsbert Erkens, University of Utrecht, The Netherlands.

5. How are changes in teachers' assessment practices reflected in their students' achievement?

By taking into account research question 4, we formulated a set of variables to be used for the first analysis of the second round of interviews where some change may already be expected.

V5 variable 1 to 4, **main codes** (alphabetic order), version 2

ccp	changes in classroom practice
cta	changes in teachers' attitude towards assessment
ss	important support from different sources

V6, **sub codes** (alphabetic order), version 2

cta – asw	analysis student work and strategies, scoring and grading, record keeping
cta – ie	instructional embedded assessment, more observations, more discussion, different role of homework
cta – lt	less tests, less quizzes, less homework checks
cta – oai	critique own assessment instruments, use of other formats
cta – pyr	more levels, assessment pyramid
cta – sr	emphasis on student responsibility
cta – und	better understanding of and more confidence in assessment issues in general

The same sub codes (except '- und') were used with main code ccp.

ss – am	AssessMath!
ss – catch	CATCH Team
ss – col	colleagues, amongst them lead teachers, leadership team
ss – cur	curriculum materials used (for example balanced assessments)
ss – gap	GAP-book
ss – mat	background materials used, e.g. balanced assessments MiC
ss – prin	principals, administrators
ss – time	release time
ss – utr	Utrecht seminar and info about assessment pyramid
ss – web	CATCH website
ss – work	workshops, summer institutes, conferences

V7, **sub code**

ns	non-success

sa	success student achievement
Furthermore sub codes to compare the answers of the second interview round with those of the first and third were used here.	

Figure 1, codes

After the initial analysis using the codes we found the interrater reliability was not sufficient, Cohen's Kappa of 0.4 was too low. In a small group codes were discussed and adapted. The second version of the codes used is shown in Figure 1. Furthermore a new column, V7 was added for coding (non) success and records that are important for comparison with earlier surveys/interviews. Codes were discussed until agreement was reached between two members of the research team, before importing them in the files.These codes will be used for the analysis of all three rounds of interviews. In the table of figure 2, results from the second round of interviews are shown:

variables (V6)	*abs.freq.*	*valid perc.*		*abs.freq.*	*valid perc.*
cta-asw	11	4.44%	**ss-am**	1	0.40%
cta-ie	18	7.26%	**ss-catch**	19	7.66%
cta-lt	1	0.40%	**ss-col**	27	10.89%
cta-oai	13	5.24%	**ss-cur**	5	2.02%
cta-pyr	20	8.06%	**ss-gap**	4	1.61%
cta-sr	1	0.40%	**ss-mat**	6	2.42%
cta-und	5	2.02%	**ss-prin**	3	1.21%
ccp-asw	18	7.26%	**ss-time**	3	1.21%
ccp-ie	37	14.92%	**ss-utr**	5	2.02%
ccp-lt	5	2.02%	**ss-web**	5	2.02%
ccp-oai	18	7.26%	**ss-work**	10	4.03%
ccp-pyr	5	2.02%	other/no code	2236	
ccp-sr	8	3.23%	Total	2484	100%

Figure 2, first results

Some remarks can be made by looking at these results:

1. The pyramid model (as one teacher noted in the initial survey: "Sweet, short and simple") proved to be an important model to initiate changes in teachers' ideas about assessment practices. However, whereas the pyramid was mentioned 20 times in relation to change in teachers' attitude towards assessment, only 5 times teachers reported a change in classroom practice in relation to the pyramid. Apparently this model is very appealing to teachers but not easy to put into practice. This is in contrast with the frequencies found in the coding of change towards instructional embedded assessment: Change in attitude towards instructional embedded assessment was mentioned 18 times, whereas 37 times a change in practice towards instructional embedded assessment was reported. Our conclusion is that this type of change is much easier for teachers to implement than the ideas related to the pyramid.

2. Often actual changes in classroom practice were mentioned. 37 times (14.9%) instructional assessment was mentioned, 18 times (7.3%) the use of other formats for assessments, 18 times (7.3%) the analysis of student work and keeping record of informal assessment and 5 times (2.0%) teachers stated the number of tests changed (quality for quantity).
3. Responsibility for their own work is a key word for teachers as well as for students. Teachers state they have a better understanding now of assessment issues (5 times, 2.0%) and some either say they want their students to have a greater responsibility (8 times, 3.2%) or that they have already achieved that (1, 0.4%).
4. The support given by the CATCH team (19 times, 7.7%) and colleagues (including the CATCH leadership team) proved to be important (27 times, 10.9%). Building a community of collaborating teachers was one of the goals of the CATCH program; these results show that goal was reached to a great extend.
5. Some teachers report they have not been successful (yet) in implementing CATCH ideas. They mention different reasons why, in their situation, or with their students, changes are impossible.

During our presentation at PME we expect to present more results since the results of the first and third round of interviews will be available by then. We will discuss our findings with the attendants.

CONCLUSIONS

Experiences in the CATCH project show that teachers as well as their students can profit from an enhanced insight in formative assessment practices. The pyramid model, as presented in the CATCH professional development program designed to support changes in teacher's formative assessment practices, proved to be an important model to change teachers' attitude towards "teaching, learning and assessing for understanding". To put these ideas into practice, however, was easier for instructional embedded assessment compared to their use in the design of teacher made tests. The ideas on which the project was based "traveled" faster and are now used by a much larger group of teachers than was expected. Issues that need further exploration are (a) the impact on student achievement on standardized tests, (b) the support offered by school personnel and (c) the way individual teachers make decisions about the assessment instruments to use, when they use them and which reasons motivate their choice.

References:

Black, P. & William, D. (1998). Inside the black box. *Phi Delta Kappa, 80*(2), 139-148

Boertien, H. & De Lange, J. (1994). *The national Option of TIMSS in The Netherlands*. Enschede: University of Twente

Bransford, J.D., Brown, A.L. & Cocking, R.R. (Eds.). (1999). *How people learn:Brain, mind, experience and school*. Washington, DC: National Academy Press

Cappo, M., de Lange, J. & Romberg, T. A. (1999). *AssessMath!* Santa Cruz, CA: Learning in Motion

Dekker, T. & Querelle, N.(2002). *Great Assessment Problems*. Utrecht, The Netherlands: Freudenthal Institute, www.fi.uu.nl/catch

De Lange, J. (1987). *Mathematics, Insight and Meaning*. Doctoral thesis. Utrecht: OW&OC

De Lange, J. (1995). *Assessment: No change without Problems*. In T.A. Romberg (Ed.), *Reform in School Mathematics* (pp. 87-172). Albany, NY: SUNY Press

De Lange, J. (1999). *Framework for classroom assessment in mathematics*. Freudenthal Institute, University of Utrecht, The Netherlands & NCISLA, University of Wisconsin, Madison

Kuiper, W.A.J.M., Bos, K. & Plomp, Tj. (2000). The TIMSS national option test. *Studies in Educational Evaluation,* 5(2), 85-104

Organisation for Economic Co-operation and Development. (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD

Romberg, T. (1999). *Monitoring student progress in mathematics*. Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science

Van den Heuvel – Panhuizen, M., Buys, K. and Treffers, A. (ed) (2000). *Kinderen leren rekenen. Tussendoelen Annex Leerlijnen. Hele Getallen Bovenbouw Basisschool*. Groningen: Wolters Noordhoff

Verhage, H.& De Lange, J. (1997). Mathematics education and assessment. *Pythagoras, 42*, 14-20

Webb, D. and Dekker, T. (2002). Classroom assessment as a basis for teacher change (CATCH). In Mewborn, D. et all (eds) *Proceedings of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. II, pp. 599 – 609, Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education